Defectives:	0	1	2
o_i :	110	75	15
e_i :	128	64	8

Test

$$H_0 : X \sim Bin(n = 2, p = 0.2)$$
$$H_a : X \nsim Bin(n = 2, p = 0.2)$$

at the $\alpha = 0.05$ significance level.

The expected counts are computed by noting that under H_0 (i.e. $X \sim Bin(n = 2, p = 0.2)$),

$$P(X = 0) = \binom{n}{x} p^x (1 - p)^{n-x} = \binom{2}{0} 0.2^0 (1 - 0.2)^{2-0} = 0.64,$$

$P(X = 1) = 0.32$, and $P(X = 2) = 0.04$. Hence, out of the 200 packages, we expect $0.64 \times 200 = 128$ to have 0 defectives, $0.32 \times 200 = 64$ to have 1 defective, and $0.04 \times 200 = 8$ to have 2 defectives; thus $e_1 = 128$, $e_2 = 64$, and $e_3 = 8$. The test statistic is

$$\chi^{2*} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \frac{(110 - 128)^2}{128} + \frac{(75 - 64)^2}{64} + \frac{(15 - 8)^2}{8} = 10.547$$

We have $k = 3$ "bins", so the degrees of freedom is $k - 1 = 2$. Thus, the p−value for the test is

$$p - \text{value} = P(\chi^2_{(k-1)} > \chi^{2*}) = P(\chi^2_{(2)} > 10.547) \in (0.005, 0.01)$$

The actual p−value is equal to 0.0051 (via computer). Since the p−value is less than α, reject H_0. We have evidence that the number of defective bulbs per package does not have a $Bin(n = 2, p = 0.2)$ distribution.

The R output for the above analysis is shown in **R 10.6** on page 188.

10.4 Exercises

♥ = answers are provided beginning on page 229.

10.1 ♥ A researcher sought to summarize the relationship between migraine headaches and caffeine consumption (low, medium, high). A random sample of 135 people yielded the following contingency table.

	Caffeine Consumption		
	Low (L)	Medium (M)	High (H)
Migraine (Mig)	5	8	15
No Migraine (Mig^c)	35	42	30

(a) What is the adverse outcome? What is the risk factor?

(b) Find the risk of migraines for high caffeine consumers, i.e. find $P(Mig|H)$.

(c) Find the risk of migraines for low caffeine consumers.

(d) Find the relative risk of migraines for high caffeine consumers versus low caffeine consumers. Interpret.

(e) Find the increased risk. Interpret.

(f) Find the odds of migraines for medium caffeine consumers.

(g) Find the odds of migraines for low caffeine consumers.

(h) Find the relative odds (odds ratio) of migraines for medium versus low caffeine consumers. Interpret.

10.2 A random sample of 50 musicians (M) had 10 with tinnitus (T) (tinnitus is a constant ringing in the ears), while 10 out of 100 randomly selected non-musicians (M^c) had tinnitus.

(a) What is the adverse outcome? What is the risk factor?

(b) What is the risk of tinnitus for musicians?

(c) What is the risk of tinnitus for non-musicians?

(d) What is the relative risk of tinnitus for musicians versus non-musicians? Interpret.

(e) What is the increased risk? Interpret.

(f) What is the odds of tinnitus for musicians?

(g) What is the odds of tinnitus for non-musicians?

(h) What is the relative odds (odds ratio) of tinnitus for musicians versus non-musicians? Interpret.

10.3 Out of 300 randomly selected welders (W), 33 suffer from retinal damage (R). Out of 250 randomly selected non-welders (W^c) (adults that have never used a welder), 20 suffer from retinal damage (R).

(a) What is the relative risk of retinal damage (R) for welders (W) versus non-welders (W^c)? Interpret.

(b) What is the relative odds (odds ratio) of retinal damage (R) for welders (W) versus non-welders (W^c)? Interpret.

10.4 ♥ In reference to question (10.1), suppose we wish to test H_0 : caffeine and migraines are independent versus H_a : caffeine and migraines are not independent at the $\alpha = 0.05$ significance level.

(a) Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.

(b) Approximate the p–value for the test using the chi-square table.

(c) Use the χ^2–Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p–value for the test.

(d) Based upon your analysis, is there a significant association between caffeine and migraines? Why?

10.5 A gas station wants to understand the relationship between the gender of its customers and their choice of gasoline. The following two-way table summarizes gender (male/female) and gasoline (regular/midgrade/premium) for 100 randomly selected customers.

	Regular	Midgrade	Premium
Male	15	15	25
Female	25	15	5

They want to test H_0 : gender and gasoline are independent versus H_a : gender and gasoline are not independent at the $\alpha = 0.10$ significance level.

(a) Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.

(b) Based upon your answer in 10.5(a), is there a significant association between gender and gasoline? Why?

(c) Approximate the p−value for the test using the chi-square table.

(d) Use the χ^2−Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p−value for the test.

(e) Based upon the p−value, is there a significant association between gender and gasoline? Why?

(f) Based upon the p−value, do we have moderate, strong, or very strong evidence of a significant association?

10.6 Consider the following two-way table which summarizes gender and job position (manager, non-manager) for 100 randomly selected employees at a large company.

	Male	Female
Manager	30	10
Non-Manager	30	30

A researcher wishes to test H_0 : no association between gender and position versus H_a : association between gender and position at the $\alpha = 0.01$ significance level.

(a) Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.

(b) Approximate the p−value for the test using the chi-square table.

(c) Use the χ^2−Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p−value for the test.

(d) Based upon the p−value, is there a significant association between gender and position? Why?

(e) At the $\alpha = 0.05$ significance level, is there a significant association between gender and position? Why?

10.7 In 2006, a random sample of 400 tax payers was collected. The following two-way table summarizes income level (low, medium, or high) versus being audited by the IRS. Note: if a person made less than $30,000, then they were classified as low income, middle income was classified as between $30,000 and $80,000, and high income was classified as over $80,000.

	Audited (A)	Not Audited (A^c)
Low Income (L)	5	150
Middle Income (M)	15	145
High Income (H)	15	70

(a) Determine the relative risk of being audited for high income tax payers versus low income tax payers. Interpret.

(b) We wish to test H_0 : income and audit are independent versus H_a : income and audit are not independent at the $\alpha = 0.10$ significance level. Find the p−value for this test using the chi-square table.

(c) Use the χ^2−Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p−value for the test.

(d) Based upon the p−value, do we have no evidence, moderate, strong, or very strong evidence of a significant association?

10.8 ♥ In the game *Twister*, participants spin a spinner. The spinner can stop in a red, blue, yellow, or green section. The spinner is supposed to yield an equal probability for each color (i.e the probability for each color is supposed to be 1/4). Suppose 40 spins yielded

Spin:	Red	Blue	Yellow	Green
o_i :	6	11	10	13
e_i :	10	10	10	10

Test

$$H_0 : \text{the spinner is fair}$$
$$H_a : \text{the spinner is not fair}$$

at the $\alpha = 0.05$ significance level. Note that under H_0 (the spinner is fair), we expect the number of times the spinner lands in red to be $e_1 = 40 \times 1/4 = 10$. The other colors are the same, therefore $e_1 = \cdots = e_4 = 10$.

(a) Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.

(b) Approximate the p−value for the test using the chi-square table.

(c) Use the χ^2−Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p−value for the test.

(d) Based upon the p−value, is there evidence that the spinner is not fair? Why?

10.9 The manufacturer of M&M's claims the following color breakdown: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, and 13% brown. A randomly selected bag of M&M's had 103 candies and yielded the following colors.

	blue	orange	green	yellow	red	brown
o_i :	25	22	19	17	7	13
e_i :	24.72	20.60				

Test

$$H_0 : \text{the manufacturers color breakdown is correct}$$
$$H_a : \text{the color breakdown is different than the manufacturers claim}$$

at the $\alpha = 0.05$ significance level. Under H_0 (i.e. under the manufacturers claimed color proportions), the number of blues that we expect is $e_1 = 103 \times 0.24 = 24.72$, the expected number of oranges is $103 \times 0.20 = 20.60$, etc.

(a) Determine the rest of the expected counts, $e_3, \ldots, e_6$.

(b) Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.

(c) Approximate the p–value for the test using the chi-square table.

(d) Use the χ^2–Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/chisq.html`

to precisely determine the p–value for the test.

(e) Based upon the p–value, do we have evidence that the color breakdown significantly differs from the manufactures claim? Why?